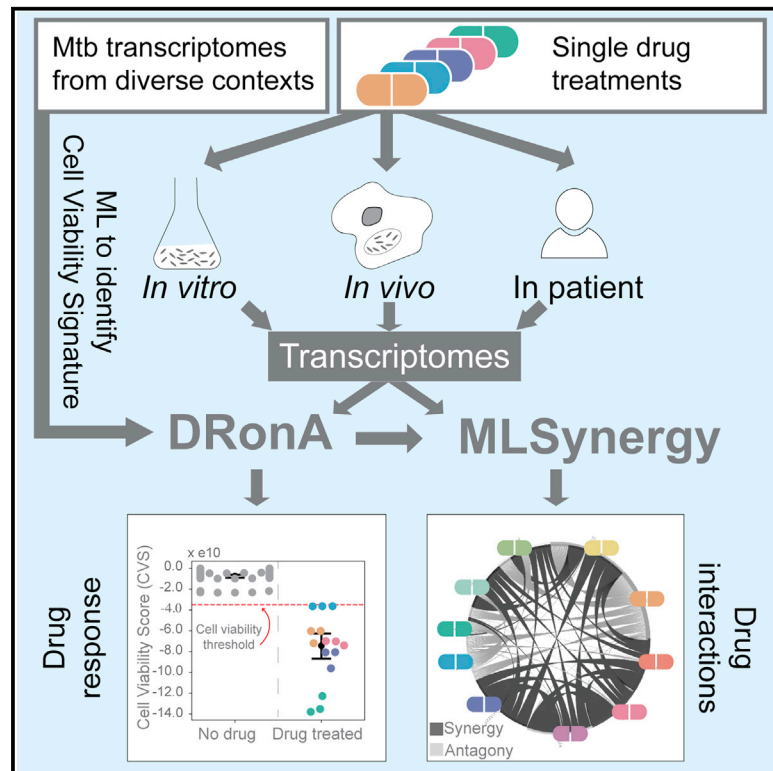**Article**

Transcriptome signature of cell viability predicts drug response and drug interaction in *Mycobacterium tuberculosis*

Graphical abstract

Authors

Vivek Srinivas, Rene A. Ruiz, Min Pan, Selva Rupa Christinal Immanuel, Eliza J.R. Peterson, Nitin S. Baliga

Correspondence

eliza.peterson@isbscience.org (E.J.R.P.), nitin.baliga@isbscience.org (N.S.B.)

In brief

To develop new, shorter tuberculosis drug regimens, there is a critical need to easily and rapidly measure treatment efficacy. Srinivas et al. report the development of DRonA and MLSynergy algorithms, which leverage transcriptomic signatures of viability for *in silico* prediction of Mtb drug response and multidrug interactions.

Highlights

- DRonA defines the transcriptome signature for *M. tuberculosis* (Mtb) viability

- DRonA detects killing of Mtb by diverse mechanisms and in diverse contexts

- DRonA predicts viability of dormant non-culturable Mtb cells

- DRonA with MLSynergy predicts the outcome of two- and three-drug combinations

CellPress

Article

Transcriptome signature of cell viability predicts drug response and drug interaction in *Mycobacterium tuberculosis*

Vivek Srinivas,[1] Rene A. Ruiz,[1] Min Pan,[1] Selva Rupa Christinal Immanuel,[1] Eliza J.R. Peterson,[1,*] and Nitin S. Baliga[1,2,3,4,5,*]

[1]Institute for Systems Biology, Seattle, WA, USA
[2]Departments of Biology and Microbiology, University of Washington, Seattle, WA, USA
[3]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA
[4]Lawrence Berkeley National Lab, Berkeley, CA, USA
[5]Lead contact
*Correspondence: eliza.peterson@isbscience.org (E.J.R.P.), nitin.baliga@isbscience.org (N.S.B.)
https://doi.org/10.1016/j.crmeth.2021.100123

MOTIVATION   The treatment of tuberculosis (TB), which kills 1.8 million each year, remains difficult, especially with the emergence of multidrug resistant strains of *Mycobacterium tuberculosis* (Mtb). While there is an urgent need for new drug regimens to treat TB, the process of drug evaluation is slow and inefficient, owing to the slow growth rate of the pathogen, the complexity of performing bacteriologic assays in a high-containment facility, and the context-dependent variability in drug sensitivity of the pathogen. The algorithms "DRonA" and "MLSynergy" described here use transcriptomes of drug-treated Mtb to predict drug response and drug interaction in diverse contexts.

SUMMARY

There is an urgent need for new drug regimens to rapidly cure tuberculosis. Here, we report the development of drug response assayer (DRonA) and "MLSynergy," algorithms to perform rapid drug response assays and predict response of *Mycobacterium tuberculosis* (Mtb) to drug combinations. Using a transcriptome signature for cell viability, DRonA detects Mtb killing by diverse mechanisms in broth culture, macrophage infection, and patient sputum, providing an efficient and more sensitive alternative to time- and resource-intensive bacteriologic assays. Further, MLSynergy builds on DRonA to predict synergistic and antagonistic multidrug combinations using transcriptomes of Mtb treated with single drugs. Together, DRonA and MLSynergy represent a generalizable framework for rapid monitoring of drug effects in host-relevant contexts and accelerate the discovery of efficacious high-order drug combinations.

INTRODUCTION

New treatment regimens containing multiple drugs are needed to achieve rapid and complete clearance of *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis (TB). However, the discovery of effective multidrug combinations is a challenging endeavor, burdened by the enormous number of testable combinations (e.g., a collection of 1,000 compounds yields ~500,000 pairwise combinations and exponentially larger numbers of higher-order combinations). Multicomponent drug discovery is particularly challenging for Mtb, a slow-growing pathogen that is capable of generating phenotypically heterogeneous subpopulations. These phenotypically diverse subpopulations allow Mtb to persist and survive the variable conditions encountered during infection as well as thwart drug treatment.

Because of drug-tolerant subpopulations within the host, a large proportion of drug regimens that are effective in killing Mtb *in vitro* are futile in patients (Ji et al., 1996; Sarathy et al., 2018, 2019). Suffice it to say, new approaches are needed to reduce the search space and prioritize combinations for experimental testing, while also taking into account the host context and different subpopulations of Mtb.

Another challenge in the development of new antitubercular drug regimens is the reliance on growth assays to monitor treatment response. Current methods to monitor treatment response include counting of colony forming units (CFUs) on solid agar plates and measuring the time it takes for a sample in liquid culture to become culture positive for Mtb, in what is termed time to positivity (TTP) assay. Both CFU counting and TTP have their drawbacks including loss of sensitivity, vulnerability to contamination,

and lengthy time to measure results. Furthermore, culture on solid media or in liquid media requires actual growth, which limits the detection of mycobacterial subpopulations that are viable but not actively growing (Diacon et al., 2012). Instead, profiling 16S ribosomal RNA as a proxy for Mtb load in sputum is emerging as a more sensitive technique that addresses the shortcomings of growth-based assays (Honeyborne et al., 2011, 2014). This is a promising development because information in RNA can be amplified using technologies such as probe capture and PCR to develop highly sensitive methods for investigating the drug response of Mtb, especially from patient samples. Furthermore, genome-wide expression studies of Mtb from broth, sputum, and *in vivo* infections have been used to uncover physiologic states and transcriptional mechanisms of drug tolerance (Honeyborne et al., 2016; Peterson et al., 2019). Here, we have investigated if transcriptome profiling of Mtb can report on the effect of drug treatment, and whether this information can also enable *in silico* identification of drug combinations that are likely to have synergistic or antagonistic effects on the pathogen.

We report the development of a framework of two algorithms drug response assayer (DRonA) and "MLSynergy" that can use transcriptomes to predict Mtb's response to drug treatment and classify two- and three-drug combinations based on the likelihood of synergistic or antagonistic action on Mtb. DRonA is a machine learning algorithm that was trained on publicly available transcriptomes of Mtb cultured in diverse conditions (with and without perturbation) to detect a gene signature for loss of Mtb viability. Using drug-induced transcriptional changes, DRonA calculates a cell viability score (CVS), which distinguishes the extent of a drug's bacteriostatic or bactericidal activity on Mtb. We demonstrate that DRonA accurately detects within the transcriptome profile of drug-treated Mtb evidence for loss of bacterial viability, regardless of the drug's mechanism of killing. Furthermore, DRonA was equally accurate in determining drug-induced viability reduction in Mtb from broth culture, macrophage infection, and patient sputum. Finally, MLSynergy uses transcriptomes from single-drug treatment to predict the interaction of drugs in combination. Using the ratio of the expected CVS (based on CVSs of individual drugs) and the predicted CVS (based on the inferred multidrug transcriptome) calculated by DRonA, MLSynergy can distinguish between synergistic and antagonistic combinations. We demonstrate that MLSynergy accurately classified experimentally determined synergistic and antagonistic combinations. Thus, the DRonA/MLSynergy framework can accelerate antitubercular drug discovery by reducing the reliance on growth-based treatment response assays and guiding the experimental assessment of novel drug combinations.

RESULTS

DRonA detects signatures for loss of viability within transcriptomes of Mtb, irrespective of mechanism of killing

To investigate whether Mtb viability can be deciphered from its transcriptome state, we sought to define a classifier that could accurately identify transcriptomes of viable Mtb. We hypothesized that the degree of deviation of a transcriptome from the

boundary defined by the above-mentioned classifier would indicate the loss of viability of Mtb cells. Further, we hypothesized that the loss of viability would be agnostic of the inhibitory effect, making it possible to predict drug-mediated killing, irrespective of the mechanism of action (Figure 1). While there are many good classification techniques (e.g., artificial neural networks, decision trees, Bayesian classifiers), the support vector machine (SVM) algorithm is one of the best techniques to optimize the expected solution (i.e., identifying a signature of viable states of Mtb) with limited datasets. Moreover, classification based on SVM offers potential for feature analysis to identify specific genes whose expression levels are diagnostic of the viability state of Mtb. Therefore, we trained a single-class support vector machine (SC-SVM) using a compendium of 3,151 transcriptomes of Mtb grown in diverse conditions to accurately identify the transcriptomes that belong to "viable" states of Mtb. The compendium of 3,151 transcriptomes was compiled from 173 studies available in the Gene Expression Omnibus (GEO). These studies used microarray and RNA sequencing (RNA-seq) to assess gene expression changes in Mtb from various growth medium compositions, culture conditions, and drug treatment (Table S1). Batch effects and platform-specific bias across the transcriptome profiles were corrected with rank normalization, and each profile was labeled as "viable," "non-viable," or "unclassified" by manual inspection of the associated metadata. Specifically, 24 transcriptomes of Mtb cultured in optimal growth conditions (mid-log phase of growth in 7H9 nutrient-rich media, incubated at 37°C with aeration) were labeled as "viable," and 193 transcriptomes of Mtb cultures treated with 17 different drugs at >1× MIC_{50} for >12 h were labeled as "non-viable". The remaining 2,319 transcriptomes were labeled as "unclassified." The labeled transcriptome compendium was used for SC-SVM training, which was performed to broaden the classifier-defined boundary of viability by iteratively including transcriptomes from the "unclassified" set that are from viable Mtb adapted to non-lethal, sub-optimal growth conditions (see STAR Methods for details). The classifier was iteratively trained on the "viable" set until addition of transcriptomes from the "unclassified" set caused a drop in its performance in accurately classifying viable and non-viable transcriptomes (Figure S1). The final classifier was trained on 994 transcriptomes (Table S1) of Mtb from diverse growth conditions, including log phase, vehicular control samples, nutrient starvation, low pH, hypoxia, and intracellular growth. As such, the SC-SVM classifier identified Mtb transcriptomes from slow-growing (i.e., dormancy-inducing), but viable conditions. In contrast, the excluded transcriptomes (total 1,940) (Table S1) were from stressful conditions (e.g., drug treated, heat treated, amino acid starved) and lethal genetic perturbations (e.g., *phoP*, *espR*, *mihF* mutants) that reduced cell viability in Mtb cultures (Foddai et al., 2010; Odermatt et al., 2018; Pérez et al., 2001; Roy et al., 2020; Tiwari et al., 2018).

The linear SC-SVM classifier, named **d**rug **r**esp**on**se **a**ssayer (DRonA), can take as input transcriptomes of Mtb to calculate a CVS (see STAR Methods for details). The calculated CVS is proportional to the deviation of a given transcriptome from the lower limit of the classifier-defined viable transcriptome space. This lower limit is set as the cell viability threshold (cell viability
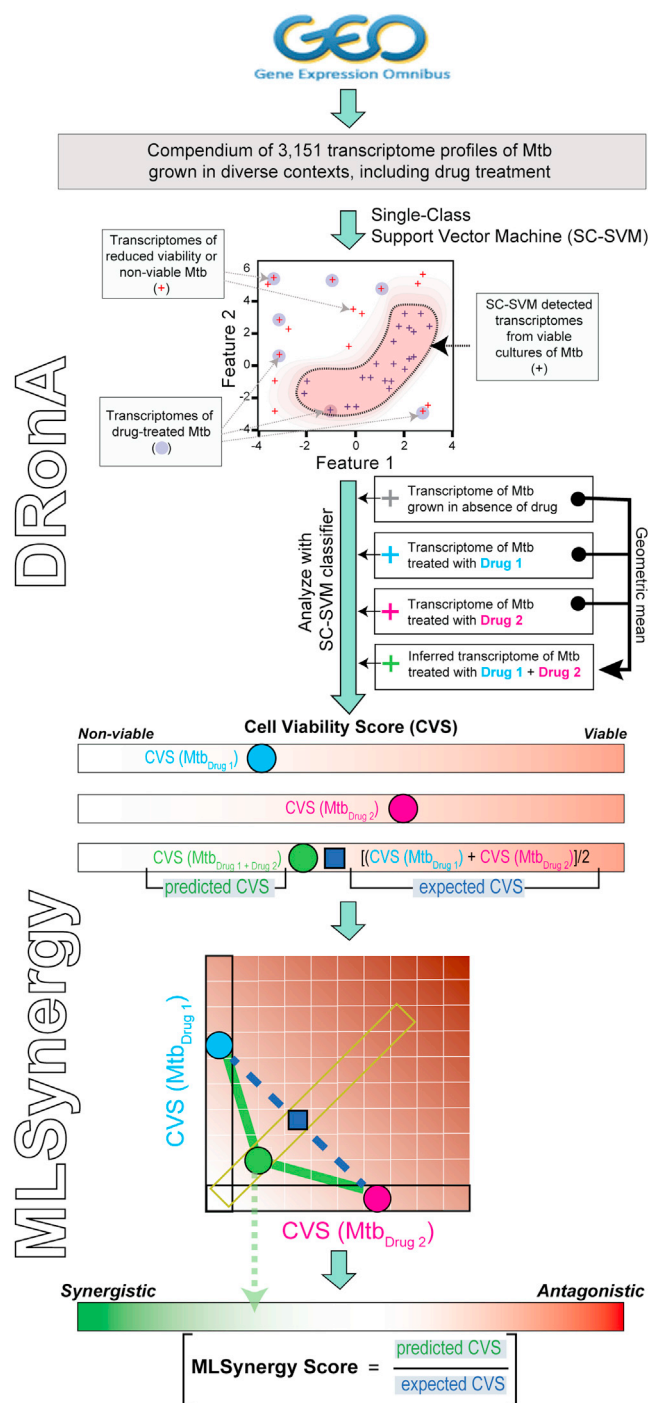
Figure 1. Overview schematic of DRonA/MLSynergy framework

DRonA is a SC-SVM that was trained on transcriptomes from viable Mtb cultures. DRonA was trained through an iterative process to define a region in the hyperplane that classifies transcriptomes from Mtb grown in varying growth conditions as viable and distinguishes them from non-viable transcriptomes (i.e., drug treated at >MIC$_{50}$ concentration). DRonA takes transcriptomes as input and outputs a CVS, which is the empirical distance from the viable class and indicative of efficacious drug treatment. Using an inferred transcriptome of a drug combination from single-drug transcriptomes, MLSynergy predicts the outcome of the drug interaction.

threshold = $-3.5e^{10}$), below which a CVS indicates a transcriptome signature of nonviable Mtb. Using an independent compendium of 72 transcriptomes generated for this study (Table S2), we ascertained that the CVS scoring scheme of DRonA accurately classified as "viable" (i.e., with a CVS $> -3.5e^{10}$) all 27 transcriptomes of Mtb grown in 7H9 medium in the absence of drugs. By contrast, DRonA predicted loss of viability (CVS $< -3.5e^{10}$) from transcriptomes of Mtb cultures treated for 72 h in 7H9 growth medium with each of the seven frontline TB drugs at $\geq$ MIC$_{50}$ concentration (p value < 0.001, Figure 2A). As expected, pyrazinamide treatment at 3.0 mg/mL was not predicted to reduce the viability of Mtb (Peterson et al., 2015). Next, we tested the performance of DRonA in predicting Mtb viability within an intracellular host context, using as input 39 transcriptomes of Mtb from naive, lipopolysaccharide (LPS)-activated, and drug-treated infected macrophages of J774A.1 lineage (Table S2) (Liu et al., 2016). Again, DRonA correctly classified the transcriptomes from untreated Mtb as viable and the drug-treated transcriptomes as non-viable (Figure 2B). Moreover, DRonA detected the known increase in the intracellular efficacy of pyrazinamide (Salfinger et al., 1990; Zhang et al., 2002) and also the decreased efficacy of rifampicin (Adams et al., 2011) in killing Mtb within macrophages. DRonA also detected a loss in the viability of Mtb within interferon-gamma-activated macrophages upon LPS treatment (Fang et al., 2020). Together this demonstrates that DRonA was able to identify non-viable transcriptomes, irrespective of the context and underlying mechanism of killing (i.e., whether immune or drug induced). Finally, we tested the performance of DRonA in predicting drug response within TB patients, using as input 16 transcriptomes of Mtb from the sputum of eight patients at the start of and after 7 or 14 days of successful TB treatment with isoniazid (H), rifampicin (R), pyrazinamide (Z), and ethambutol (E) (Honeyborne et al., 2016). DRonA efficiently differentiated cell viability from the Mtb transcriptomes collected from patients on day 0 from transcriptomes collected on day 7 or 14 of drug treatment (p value < 0.01) (Figure 2C), demonstrating that DRonA can detect drug treatment response from bacterial RNA in patient sputum.

DRonA can estimate the decline in CFUs upon drug treatment

We next tested whether the CVS was proportional to the magnitude of drug effects based on CFU assessment. We compared DRonA-generated CVSs with the relative CFUs observed after Mtb was treated for 24 and 72 h with seven frontline TB drugs at various concentrations and conditions (Tables S3 and S4). The CVS scores calculated from transcriptomes of both untreated Mtb cultures and those treated with drugs at < MIC$_{50}$ concentrations were higher than the viability threshold. Although, the inferred CVS from cultures treated with <MIC$_{50}$ drug was less than the CVS of untreated cultures (difference in average = $-3.5 \times e^{10}$, p value < 0.01), indicating a moderate loss of viability. In contrast, the CVS scores calculated from transcriptomes of Mtb cultures treated with $\geq$MIC$_{50}$ concentration of drug were

---

MLSynergy uses the Loewe additivity principle and calculates the ratio of predicted CVS to expected CVS to determine synergy or antagonism for drug combinations.
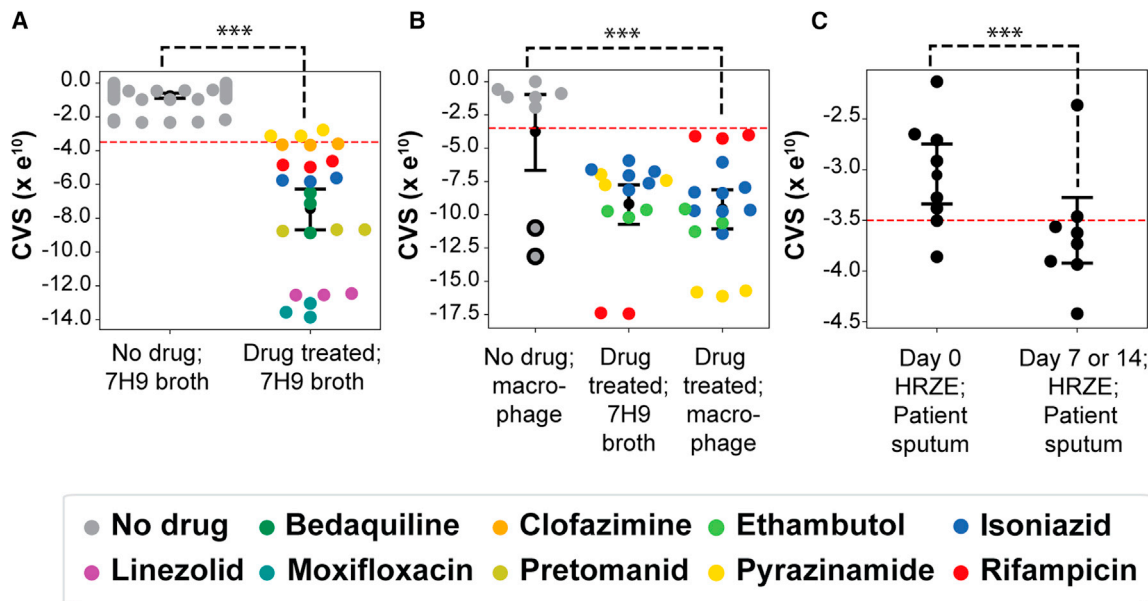
Figure 2. DRonA-generated CVSs for transcriptomes of Mtb sourced from broth culture, macrophage infection, and patient sputum
(A) CVSs for transcriptomes of Mtb cultures grown in 7H9-rich media with or without drug treatment for 72 h.
(B) CVSs for transcriptomes of Mtb cultured in 7H9 broth with drug treatment for 24 h and macrophage with or without drug treatment for 24 h. Circles with black border indicate transcriptomes from interferon-gamma-activated macrophages with lipopolysaccharide treatment.
(C) CVSs for transcriptomes of Mtb in patient sputum collected at the start and end of 7 or 14 day chemotherapy with HRZE: isoniazid (H), rifampicin (R), pyrazinamide (Z), and ethambutol (E). Red dashed line is the cell viability threshold ($-3.5e^{10}$), below which the samples are considered to be non-viable. Black dot and error bars indicate the mean and standard deviation away from the mean. Statistical significance (black dashed line) was calculated as p value with Student's t test. ***: p value < 0.001.

consistently below the viability threshold. Furthermore, for both Mtb grown in 7H9 medium and within macrophages, the reduction in CVS was proportional to the decrease in CFU for most drugs (Figures 3 and S2), with the exception of bedaquiline. It is known that bedaquiline kills Mtb relatively slowly compared with other frontline drugs and could explain the discord between CFU and CVS within 72 h of treatment (Andries et al., 2005; Koul et al., 2007, 2008, 2014; Peterson et al., 2016). It is remarkable that despite the slow bactericidal activity of bedaquiline, its lethal effect was captured in transcriptome changes at a significantly earlier time point, and we still observed a significant correlation between relative CFUs and CVS across all drug treatments ($r = -0.9$, p value < 0.001).

A notable disadvantage of performing drug response assessment via CFU counting is the limitation that it only measures culturable bacteria (Dusthackeer et al., 2019). Mtb from *in vivo* models of latent TB infection are non-culturable and require resuscitation-promoting factors or conditions to resume growth (Biketov et al., 2007; Dhillon et al., 2004). Thus, CVS scores determined using mRNA signatures represent a comprehensive assay of drug effects on dormant Mtb that are unable to grow on solid medium but retain full potential of recovering to a physiologically active state. To test this hypothesis, we investigated the accuracy of DRonA in predicting Mtb killing by a moderate concentration of rifampicin (5 μg/mL) in potassium-deficient growth medium (Ignatov et al., 2015). Mtb shifts to a dormant state that is unable to grow on solid medium, but able to recover and proliferate in albumin, dextrose, and sodium chloride (ADC)-supplemented Sauton

medium containing potassium (Salina et al., 2019). The results demonstrated that CFU counting overestimated rifampicin-treatment-induced killing of the pathogen, as demonstrated by a minimum probable number (MPN) performed in the same context in ADC-supplemented liquid Sauton medium. Notably, similar to MPN results, there was no significant drop in CVS, demonstrating that DRonA accurately predicted the overall drug response in cultures that consist of non-culturable Mtb (Figure 4).

MLSynergy accurately predicts synergistic and antagonistic drug combinations from transcriptomes of single-drug-treated Mtb
Given that DRonA can detect Mtb's response to drug treatment from gene expression data, we investigated if DRonA could be used to accelerate multicomponent drug discovery by predicting the outcome of drug interactions from single-drug-treated transcriptomes. To do this, we developed an approach to infer the transcriptomes of multidrug treatments. Specifically, we inferred the transcriptome of multidrug combinations by triangulation (see STAR Methods for details) of the respective transcriptomes obtained from single-drug-treated cultures of Mtb and then used the inferred multidrug transcriptome with DRonA to predict the CVS of the multidrug combination (i.e., the "predicted CVS"). Transcriptomes used for prediction of drug interactions were from Mtb treated with single drugs in matched experimental conditions (7H9 medium and 72 h drug treatment). Using this method to predict the CVS of multidrug combinations, we developed a parametric method, "MLSynergy", to predict the interaction outcome
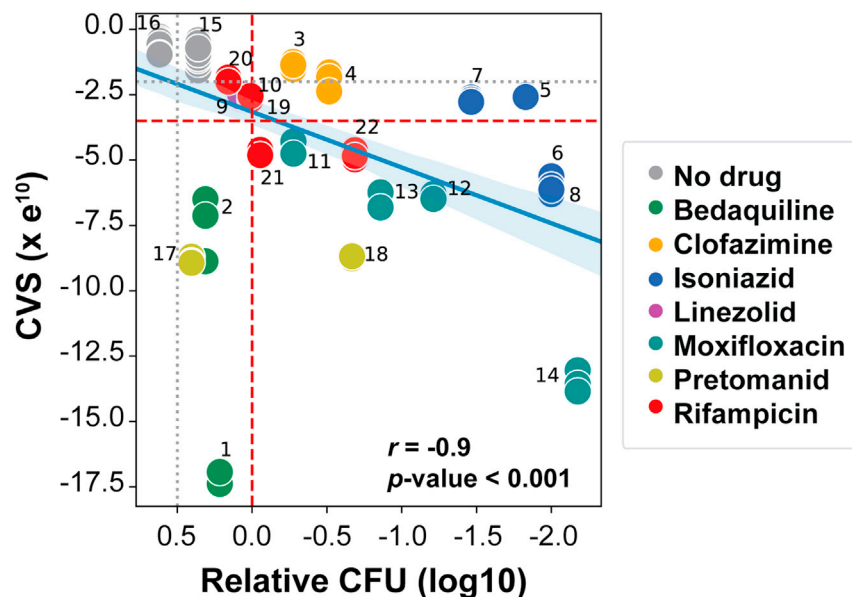
Figure 3. Correlation between CVS and relative CFU with and without drug treatment

Relative CFU was calculated in relation to 0 h (prior to drug or vehicle control treatment). Numbers associated with the points indicate specific drug treatment time and concentrations found in Table S3. Relative CFUs for the treatments in Table S3 were calculated with time-kill assay and are given in Table S4. Solid blue line denotes the Pearson's correlation between CVS and relative CFU. Significance was calculated as the average correlation coefficient, r, from 100 iterations performed with 70% randomly selected data. Shaded blue region denotes the variation in r from 100 iterations. Black dotted line denotes 50% growth inhibition from drug treatments and its corresponding CVS threshold ($-2.25e^{10}$). Dashed red line indicates bactericidal activity and its corresponding CVS threshold ($-3.5e^{10}$).

of the two- and three-drug combinations. MLSynergy predicts the synergy or antagonism of multidrug combinations based on the Loewe additivity principle (Loewe, 1928) by calculating the ratio of predicted CVS to expected CVS, where the "expected CVS" for a drug combination is the average of CVSs from respective single-drug treatments (Figure 1). For example, the predicted CVS of the antagonistic combination linezolid and moxifloxacin (Cokol et al., 2017; Deshpande et al., 2016) is greater than the expected CVS and lies above the additive plane (Figure 5A), whereas the predicted CVS of the synergistic combination linezolid and POA is less than the expected CVS and lies below the additive plane (Figure 5B). Finally, the predicted CVS of linezolid with itself is the same as the expected CVS and lies on the additive plane, consistent with the Loewe additivity principle, which states that a drug in combination with itself is additive in interaction (Loewe, 1928) (Figure 5C). As such, an MLSynergy score <1 predicts that the drug interaction is synergistic, and a score >1 indicates an antagonistic drug interaction. We calculated MLSynergy scores for all two- and three-drug combinations of eight frontline drugs (Table S2). We compared MLSynergy predictions of 26 two-drug and 40 three-drug combinations of the eight frontline drugs with their experimentally determined interaction, quantified by fractional inhibitory concentrations (FICs) (Larkins-Ford et al., 2021). This comparative analysis demonstrated that MLSynergy was >90% accurate in predicting synergistic and antagonistic effects of two- and three-drug combinations (Figure 5D and 4E). Moreover, MLSynergy scores were highly correlated with the FIC values (ρ = 0.61, p value < 0.001, Figure S3). Interestingly, three two-drug combinations (identified with red font in Figure 5D) were predicted as synergistic by MLSynergy, but were determined to be antagonistic by DiaMOND assay (Larkins-Ford et al., 2021). Notably, these combinations were previously determined to be synergistic by other studies (Cokol et al., 2017; Diacon et al., 2012; Maltempe et al., 2017), suggesting that the effect of their drug interaction could be highly dependent on the

assay method and conditions. Finally, we checked the ability of MLSynergy to predict condition-dependent drug interactions in Mtb, using as input 22 transcriptomes of Mtb from untreated and drug-treated infected macrophages of J774A.1 lineage (Table S2) (Liu et al., 2016). We predicted drug interaction for two- and three-drug combinations of isoniazid, rifampicin, and pyrazinamide in both broth culture and macrophages and compared the MLSynergy predictions with their experimental FIC values (Larkins-Ford et al., 2021) (Table S5). MLSynergy predicted that all the drug combinations are synergistic in 7H9 media and turn antagonistic in macrophage. Similarly, the experimental results found that mostly all the drug combinations (with the exception of isoniazid + rifampicin) are synergistic in broth and antagonistic in macrophage. This demonstrates that MLSynergy is robust to the context in which a drug effect is measured, and it can predict condition-dependent drug interactions. The benefit of MLSynergy to perform condition-dependent predictions will become clearer upon further availability of drug-treated transcriptomes of Mtb from macrophages and other *in vivo* contexts.

DISCUSSION

Here, we report the first machine learning framework for drug response prediction in Mtb. DRonA enables efficient prediction of cell viability from transcriptomic signatures of perturbation, including drug treatment. Using DRonA estimates of cell viability from single-drug transcriptomic data, MLSynergy can then predict synergy and antagonism of antitubercular drug combinations. Our analysis using DRonA found a strong association between *in silico* estimates of cell viability following drug treatment and experimentally observed reduction in CFUs. Moreover, the loss of viability captured by DRonA from Mtb transcriptomes of patients undergoing HRZE treatment supports the clinical utility of our approach. Finally, we report several synergistic drug combinations, suggesting that the DRonA/MLSynergy framework is a promising tool for the prioritization of new multicomponent drug regimens. While we validated predictions of two- and three-drug
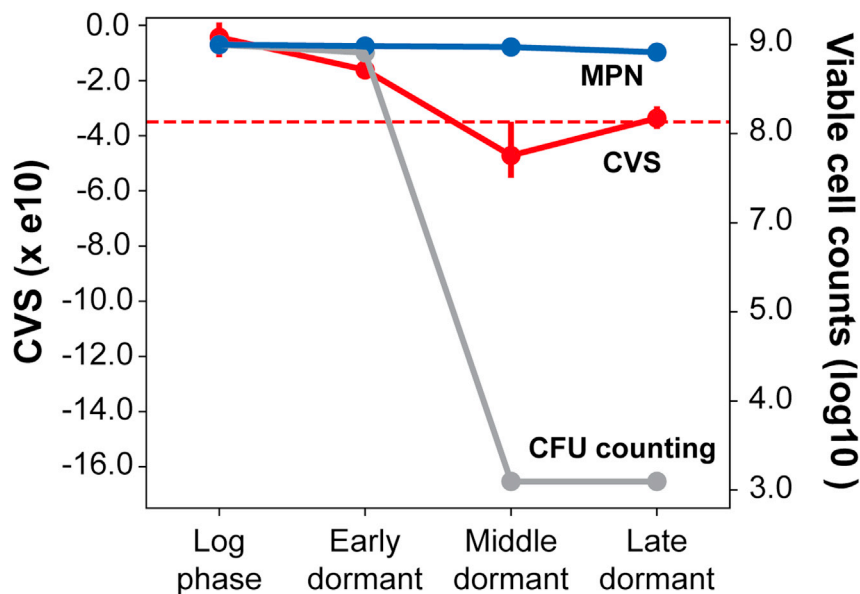
Figure 4. Comparison of CVS from DRonA with bacteriological assays determined by most probable number (MPN) assay and colony forming unit (CFU) enumeration from heterogeneous K+ starved cultures of Mtb

Transcriptomes for DRonA prediction were obtained from GEO accession number GSE66408 and viable cell counts according to MPN and CFU counting assays were obtained from Ignatov et al. (2015). Log phase is the exponentially growing cultures of Mtb collected prior to K+ starvation, and early, middle, and late dormant are the rifampicin (5 µg/mL)-treated cultures collected after 10, 20, and 30 days of K+ starvation. Red, blue and grey dots indicate mean, and error bars indicate standard deviation from the mean. The dashed red line is the CVS threshold ($-3.5e^{10}$) from DRonA and indicates loss of cell viability.

interactions, our framework is generalizable for higher-order combinations.

The suitability of using the transcriptome as a reflection of Mtb viability was studied by treating Mtb with seven frontline drugs and isolating RNA for transcriptome profiling, while also evaluating cell viability by CFU. The DRonA predicted the CVS of Mtb exposed to bactericidal (i.e., >MIC$_{50}$) concentrations of drugs were below the cell viability threshold, proportional to relative CFU and significantly different from the CVS of untreated Mtb cultured for the same duration as drug treatment. Moreover, DRonA was able to perform effectively with other transcriptomic datasets of Mtb drug treatment, including during macrophage infection and from TB patients. The ability of DRonA to accurately predict the consequence of drug treatment in 7H9 medium, within macrophages, and from patient sputum, demonstrates that the definitions of viability in the DRonA model are inclusive of both actively dividing and slow replicating (physiologically adapted) phenotypes of Mtb. Moreover, the accuracy across datasets offers DRonA as a generalizable tool for use across drug response screens and in studies where gene expression was analyzed, but Mtb viability was not measured.

Here, we showed that DRonA complements bacteriological assays in evaluating treatment response. The decline in CVS corresponded to the decline in the proportions of surviving bacilli upon drug treatment, as measured by the relative CFU counts. Since no culturing is required, DRonA can estimate drug effects much faster than conventional bacteriological assays. Additionally, the ability to enrich and amplify RNA should allow DRonA to be used with samples where bacterial cell numbers are low (Mothershed and Whitney, 2006). The high sensitivity and the autonomy from culturing makes DRonA especially promising to evaluate the efficacy of treatment regimens on dormant non-culturable Mtb that are associated with latent infection in humans.

Using DRonA-predicted viability scores, MLSynergy accurately predicted synergy and antagonism for two- and three-drug combi-

nations. This performance compares with INDIGO-MTB (Ma et al., 2019), an existing strategy that quantifies synergistic and antagonistic drug regimens using transcriptomes of Mtb treated with individual drugs, but only with drugs with known drug-drug interactions. INDIGO-MTB requires known drug-drug interactions to learn patterns and identify combinations most likely to be synergistic. In contrast, the DRonA/MLSynergy platform is based on gene signatures of cell viability and does not require any input data related to drug combinations. Comparing the accuracy for drugs without prior drug interaction information, MLSynergy significantly outperforms INDIGO-MTB (p value > 0.05, Figure S4). As such, our models can be more easily applied (i.e., no re-training required) to predict drug interaction for new drugs and conditions.

The DRonA/MLSynergy platform does have some potential limitations. First, our approach to predict cell viability based on gene expression signatures does not reveal information about drug mechanism of action. As cell death can lead to transcriptomic changes unrelated to drug treatment, these cell death signatures could be a confounding factor to make inferences about drug mechanism of action. Removing genes highly correlated to cell death could improve mechanism of action identification (Szalai et al., 2019). Future work also aims to integrate DRonA with regulatory-metabolic networks (Chandrasekaran and Price, 2010; Immanuel et al., 2021; Peterson et al., 2014; Turkarslan et al., 2015) to reveal the underlying pathways that contribute to a drug's bactericidal activity and interaction with other drugs.

Second, the DRonA/MLSynergy platform requires transcriptome profiling of Mtb drug treatment to predict drug interactions. However, we argue that predicting drug interactions using transcriptome analysis with DRonA/MLSynergy is cheaper and faster, as compared with bacteriological assays. Evaluating drug interactions with bacteriological assays requires a significantly larger number of experiments, which increases exponentially with every new drug and for testing higher-order (i.e., three-drug) interactions. For example, to evaluate all possible two-drug interactions between 10 drugs (i.e., 45 combinations), a checkerboard or DiaMOND assay would require a minimum of 55 experiments, whereas MLSynergy would require just 10 experiments to generate
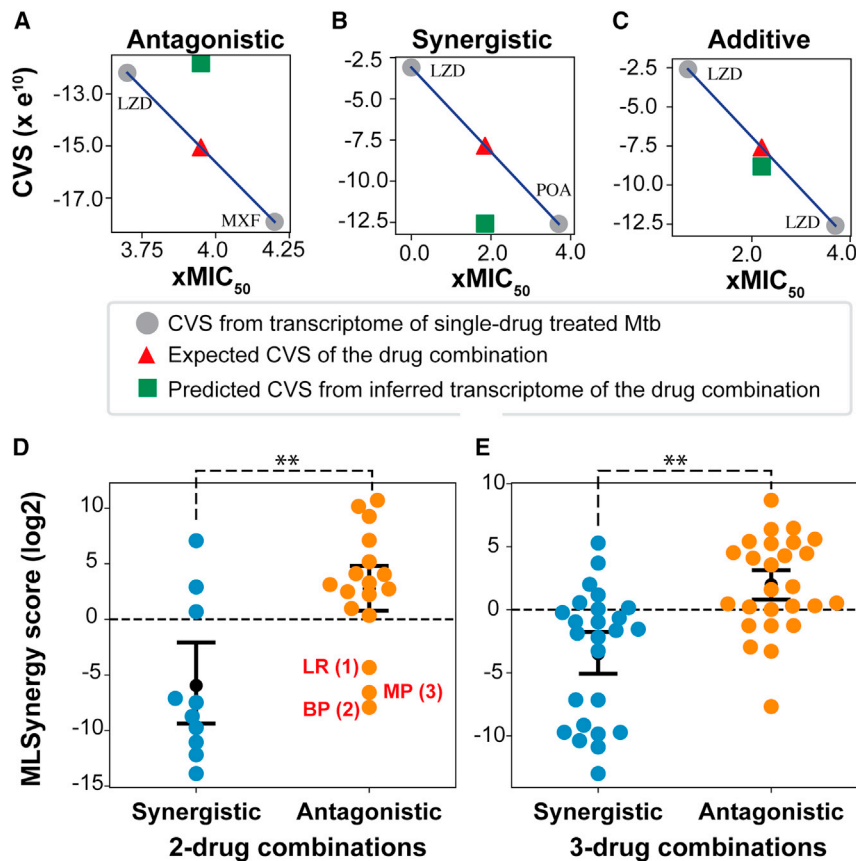
Figure 5. MLSynergy prediction of drug interaction

(A–C) Examples of the relationship between expected CVS and predicted CVS for (A) antagonistic, (B) synergistic, and (C) additive drug combinations. The expected CVS (red triangle) was calculated as the average of DRonA-generated CVSs for experimentally measured transcriptomes from single-drug-treated Mtb. LZD; linezolid, POA; pyrazinoic acid, MXF; moxifloxacin.

(D) MLSynergy classification of experimentally validated synergistic and antagonistic two-drug combinations. Drug combinations labeled in red: (1) LR; linezolid and rifampicin (Maltempe et al., 2017), (2) BP; bedaquiline and pretomanid (Cokol et al., 2017), and (3) MP; moxifloxacin and pretomanid (de Miranda Silva et al., 2019) were classified as synergistic by MLSynergy and experiments from literature, but antagonistic by Larkins-Ford et al. (2021). (E) MLSynergy classification of experimentally validated synergistic and antagonistic three-drug combinations. Black dot and error bars indicate the mean and standard deviation away from the mean. Statistical significance (black dashed line) was calculated as p value with Student's T test. **: p value < 0.01.

transcriptomes of Mtb in response to treatment with each of the 10 drugs. For three-drug combinations, checkerboard or DiaMOND assay requirement increases to 120 drug dose titration experiments, whereas requirements for MLSynergy remains the same (i.e., 10 experiments). Furthermore, technological advancements are making it faster and cheaper to profile the transcriptome of Mtb directly from patient samples (Peterson et al., 2019), which could potentially extend the utility of DRonA in rapid point-of-care devices for evaluating the effectiveness of drug treatment in TB patients.

Drug response prediction with machine learning models is an important area of current research, particularly for a slow-growing pathogen, and our results highlight the practicality of using transcriptome signatures to address major bottlenecks in the drug discovery process. The ability to detect changes in cell viability and predict drug interaction using just transcriptome profiles could substantially accelerate TB drug discovery efforts. Recent studies have demonstrated that efficacy of the same drug combination can vary significantly between broth conditions and animal models (Larkins-Ford et al., 2021). DRonA and MLSynergy could be valuable for prioritizing drug combinations that are likely to be effective in animal models, given the challenges in performing high-throughput drug assays in mouse models and non-human primates. Finally, the DronA/MLSynergy framework can be easily extended to predict other genotypes and phenotypes of Mtb associated with a gain in drug resistance (e.g., metabolic states and cell wall

composition), which could further improve treatment response prediction and clinical outcomes. To facilitate the widespread usage of these resources, the compendia of Mtb transcriptomes collated from GEO and generated in this study are publicly available, together with DRonA and MLSynergy algorithms and models in the MTB Network Portal (http://networks.systemsbiology.net/mtb/).

Limitations of the study

While we have demonstrated that DRonA is effective in quantifying loss of viability of sensitive and resistant strains of Mtb in patient sputa, and even in mixed cultures consisting of dormant bacteria, we are yet to demonstrate the utility of this technique in quantifying persister load. Persisters of Mtb can survive drug pressure and cause treatment failure, and therefore, detecting and quantifying persisters is essential to ascertain complete bacterial clearance. Time-kill curves with CFU counting has been the gold standard for quantifying persisters at single-cell resolution. In principle, as and when technologies to profile Mtb transcriptomes at single-cell resolution are sufficiently mature, DRonA could replace CFU counting for accurate quantification of persisters from host-relevant conditions, including granulomas. Similarly, the DRonA/MLSynergy platform can predict whether drugs would combine synergistically or antagonistically in any given context from which Mtb transcriptomes can be derived. This capability is important, because the nature of drug interactions can vary dramatically from broth to *in vivo* conditions. The ability of generating sufficient numbers of transcriptome profiles of drug-treated Mtb from an *in vivo* context, such as from granuloma or from caseum, remains a key limitation in demonstrating the utility of DRonA and

MLSynergy capability to predict drug interactions in an *in vivo* context. We could address this limitation by leveraging methods such as Path-seq to enrich and quantify Mtb transcriptomes from *in vivo* contexts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains and growth conditions
- METHOD DETAILS
  - Minimum inhibitory concentration 50 (MIC_{50}) determination
  - Time-kill assays
  - Collection, RNA extraction, and analysis of single-drug transcriptomes
  - Collection and curation of Mtb-GEO dataset for training of DRonA
  - Manual labeling of Mtb transcriptomes
  - Training and running DRonA
  - Inference of multi-drug transcriptomes (triangulation)
  - Calculation of MLSynergy scores for drug combinations
  - Comparison of INDIGO-MTB and MLSynergy predictions
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2021.100123.

AUTHOR CONTRIBUTIONS

V.S., E.J.R.P., and N.S.B. conceptualized the study. V.S., R.A.R., M.P., S.R.C.I., E.J.R.P., and N.S.B. designed research. V.S., R.A.R., M.P., and E.J.R.P. performed drug treatment assays. V.S., S.R.C.I., and E.J.R.P. analyzed data and performed computational analyses. V.S., E.J.R.P., and N.S.B wrote the paper. E.J.R.P. and N.S.B. acquired funding and provided supervision.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adams, K.N., Takaki, K., Connolly, L.E., Wiedenhoft, H., Winglee, K., Humbert, O., Edelstein, P.H., Cosma, C.L., and Ramakrishnan, L. (2011). Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. Cell *145*, 39–53.

Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H.W.H., Neefs, J.-M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., et al. (2005). A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science *307*, 223–227.

Biketov, S., Potapov, V., Ganina, E., Downing, K., Kana, B.D., and Kaprelyants, A. (2007). The role of resuscitation promoting factors in pathogenesis and reactivation of *Mycobacterium tuberculosis* during intra-peritoneal infection in mice. BMC Infect. Dis. *7*, 146.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Chandrasekaran, S., and Price, N.D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and *Mycobacterium tuberculosis*. PNAS *107*, 17845–17850.

Cokol, M., Kuru, N., Bicak, E., Larkins-Ford, J., and Aldridge, B.B. (2017). Efficient measurement and factorization of high-order drug interactions in *Mycobacterium tuberculosis*. Sci. Adv. *3*, 1701881.

Deshpande, D., Srivastava, S., Nuermberger, E., Pasipanodya, J.G., Swaminathan, S., and Gumbo, T. (2016). Concentration-dependent synergy and antagonism of linezolid and moxifloxacin in the treatment of childhood tuberculosis: the dynamic duo. Clin. Infect. Dis. *63*, S88–S94.

Dhillon, J., Lowrie, D.B., and Mitchison, D.A. (2004). *Mycobacterium tuberculosis* from chronic murine infections that grows in liquid but not on solid medium. BMC Infect. Dis. *4*, 51.

Diacon, A.H., Dawson, R., von Groote-Bidlingmaier, F., Symons, G., Venter, A., Donald, P.R., van Niekerk, C., Everitt, D., Winter, H., Becker, P., et al. (2012). 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. Lancet *380*, 986–993.

Dusthackeer, A., Balasubramanian, M., Shanmugam, G., Priya, S., Nirmal, C.R., Sam Ebenezer, R., Balasubramanian, A., Mondal, R.K., Thiruvenkadam, K., Hemanth Kumar, A.K., et al. (2019). Differential culturability of *Mycobacterium tuberculosis* in culture-negative sputum of patients with pulmonary tuberculosis and in a simulated model of dormancy. Front. Microbiol. *10*, 2381.

Fang, F., Ge, Q., Li, R., Lv, J., Zhang, Y., Feng, A., Kelly, G.T., Wang, H., Wang, X., Song, C., et al. (2020). LPS restores protective immunity in macrophages against *Mycobacterium tuberculosis* via autophagy. Mol. Immunol. *124*, 18–24.

Foddai, A., Elliott, C.T., and Grant, I.R. (2010). Rapid assessment of the viability of *Mycobacterium avium* subsp. paratuberculosis cells after heat treatment, using an optimized phage amplification assay. Appl. Environ. Microbiol. *76*, 1777–1782.

Honeyborne, I., McHugh, T.D., Phillips, P.P.J., Bannoo, S., Bateson, A., Carroll, N., Perrin, F.M., Ronacher, K., Wright, L., van Helden, P.D., et al. (2011). Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum *Mycobacterium tuberculosis* bacillary load during treatment. J. Clin. Microbiol. *49*, 3905–3911.

Honeyborne, I., Mtafya, B., Phillips, P.P.J., Hoelscher, M., Ntinginya, E.N., Kohlenberg, A., Rachow, A., Rojas-Ponce, G., McHugh, T.D., Heinrich, N., et al. (2014). The molecular bacterial load assay replaces solid culture for

measuring early bactericidal response to antituberculosis treatment. J. Clin. Microbiol. *52*, 3064–3067.

Honeyborne, I., McHugh, T.D., Kuittinen, I., Cichonska, A., Evangelopoulos, D., Ronacher, K., van Helden, P.D., Gillespie, S.H., Fernandez-Reyes, D., Walzl, G., et al. (2016). Profiling persistent tubercule bacilli from patient sputa during therapy predicts early drug efficacy. BMC Med. *14*, 68.

Ignatov, D.V., Salina, E.G., Fursov, M.V., Skvortsov, T.A., Azhikina, T.L., and Kaprelyants, A.S. (2015). Dormant non-culturable *Mycobacterium tuberculosis* retains stable low-abundant mRNA. BMC Genomics *16*, 954.

Immanuel, S.R.C., Arrieta-Ortiz, M.L., Ruiz, R.A., Pan, M., Lomana, A.L.G.de, Peterson, E.J.R., and Baliga, N.S. (2021). Quantitative prediction of conditional vulnerabilities in regulatory and metabolic networks of *Mycobacterium tuberculosis*. BioRxiv. https://doi.org/10.1101/2021.01.29.428876.

Ji, B., Levy, L., and Grosset, J.H. (1996). Chemotherapy of leprosy: progress since the Orlando Congress, and prospects for the future. Int. J. Lepr. Other Mycobact. Dis. *64*, S80–S88, discussion S88-90.

Koul, A., Dendouga, N., Vergauwen, K., Molenberghs, B., Vranckx, L., Willebrords, R., Ristic, Z., Lill, H., Dorange, I., Guillemont, J., et al. (2007). Diarylquinolines target subunit c of mycobacterial ATP synthase. Nat. Chem. Biol. *3*, 323–324.

Koul, A., Vranckx, L., Dendouga, N., Balemans, W., Van den Wyngaert, I., Vergauwen, K., Göhlmann, H.W.H., Willebrords, R., Poncelet, A., Guillemont, J., et al. (2008). Diarylquinolines are bactericidal for dormant mycobacteria as a result of disturbed ATP homeostasis. J. Biol. Chem. *283*, 25273–25280.

Koul, A., Vranckx, L., Dhar, N., Göhlmann, H.W.H., Özdemir, E., Neefs, J.-M., Schulz, M., Lu, P., Mørtz, E., McKinney, J.D., et al. (2014). Delayed bactericidal response of *Mycobacterium tuberculosis* to bedaquiline involves remodelling of bacterial metabolism. Nat. Commun. *5*, 3369.

Larkins-Ford, J., Greenstein, T., Van, N., Degefu, Y.N., Olson, M.C., Sokolov, A., and Aldridge, B.B. (2021). Systematic measurement of combination-drug landscapes to predict in vivo treatment outcomes for tuberculosis. Cell Syst. https://doi.org/10.1016/j.cels.2021.08.004.

Liu, Y., Tan, S., Huang, L., Abramovitch, R.B., Rohde, K.H., Zimmerman, M.D., Chen, C., Dartois, V., VanderVen, B.C., and Russell, D.G. (2016). Immune activation of the host cell induces drug tolerance in *Mycobacterium tuberculosis* both *in vitro* and *in vivo*. J. Exp. Med. *213*, 809–825.

Loewe, S. (1928). Die quantitativen Probleme der Pharmakologie. Ergebnisse der Physiologie *27*, 47–187.

Ma, S., Jaipalli, S., Larkins-Ford, J., Lohmiller, J., Aldridge, B.B., Sherman, D.R., and Chandrasekaran, S. (2019). Transcriptomic signatures predict regulators of drug synergy and clinical regimen efficacy against tuberculosis. MBio *10*, e02627–19.

Maltempe, F.G., Caleffi-Ferracioli, K.R., do Amaral, R.C.R., de Oliveira Demitto, F., Siqueira, V.L.D., de Lima Scodro, R.B., Hirata, M.H., Pavan, F.R., and Cardoso, R.F. (2017). Activity of rifampicin and linezolid combination in *Mycobacterium Tuberculosis*. Tuberculosis *104*, 24–29.

de Miranda Silva, C., Hajihosseini, A., Myrick, J., Nole, J., Louie, A., Schmidt, S., and Drusano, G.L. (2019). Effect of moxifloxacin plus pretomanid against *Mycobacterium tuberculosis* in log phase, acid phase, and nonreplicating-persister phase in an in vitro assay. Antimicrob. Agents Chemother. *63*, e01695–18.

Mothershed, E.A., and Whitney, A.M. (2006). Nucleic acid-based methods for the detection of bacterial pathogens: present and future considerations for the clinical laboratory. Clin. Chim. Acta *363*, 206–220.

Odermatt, N.T., Sala, C., Benjak, A., and Cole, S.T. (2018). Essential Nucleoid associated protein mIHF (Rv1388) controls virulence and housekeeping genes in *Mycobacterium tuberculosis*. Sci. Rep. *8*, 14214.

Pérez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., and Martín, C. (2001). An essential role for phoP in *Mycobacterium tuberculosis* virulence. Mol. Microbiol. *41*, 179–187.

Peterson, E.J., Bailo, R., Rothchild, A.C., Arrieta-Ortiz, M.L., Kaur, A., Pan, M., Mai, D., Abidi, A.A., Cooper, C., Aderem, A., et al. (2019). Path-seq identifies an

essential mycolate remodeling program for mycobacterial host adaptation. Mol. Syst. Biol. *15*, 8584.

Peterson, E.J.R., Reiss, D.J., Turkarslan, S., Minch, K.J., Rustad, T., Plaisier, C.L., Longabaugh, W.J.R., Sherman, D.R., and Baliga, N.S. (2014). A high-resolution network model for global gene regulation in *Mycobacterium tuberculosis*. Nucleic Acids Res. *42*, 11291–11303.

Peterson, E.J.R., Ma, S., Sherman, D.R., and Baliga, N.S. (2016). Network analysis identifies Rv0324 and Rv0880 as regulators of bedaquiline tolerance in *Mycobacterium tuberculosis*. Nat. Microbiol. *1*, 16078. https://doi.org/10.1038/nmicrobiol.2016.78.

Peterson, E.J.R., Abidi, A.A., Arrieta-Ortiz, M.L., Aguilar, B., Yurkovich, J.T., Kaur, A., Pan, M., Srinivas, V., Shmulevich, I., and Baliga, N.S. (2020). Intricate genetic programs controlling dormancy in *Mycobacterium tuberculosis*. Cell Rep. *31*, 107577.

Peterson, N.D., Rosen, B.C., Dillon, N.A., and Baughn, A.D. (2015). Uncoupling environmental pH and intrabacterial acidification from pyrazinamide susceptibility in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. *59*, 7320–7326.

Roy, S., Ghatak, D., Das, P., and BoseDasgupta, S. (2020). ESX secretion system: the gatekeepers of mycobacterial survivability and pathogenesis. Eur. J. Microbiol. Immunol. (Bp) *10*, 202–209.

Salfinger, M., Crowle, A.J., and Reller, L.B. (1990). Pyrazinamide and pyrazinoic acid activity against tubercle bacilli in cultured human macrophages and in the BACTEC system. J. Infect. Dis. *162*, 201–207.

Salina, E.G., Grigorov, A.S., Bychenko, O.S., Skvortsova, Y.V., Mamedov, I.Z., Azhikina, T.L., and Kaprelyants, A.S. (2019). Resuscitation of dormant "nonculturable" *Mycobacterium tuberculosis* is characterized by immediate transcriptional burst. Front. Cell. Infect. Microbiol. *9*, 272.

Sarathy, J., Blanc, L., Alvarez-Cabrera, N., O'Brien, P., Dias-Freedman, I., Mina, M., Zimmerman, M., Kaya, F., Ho Liang, H.-P., Prideaux, B., et al. (2019). Fluoroquinolone efficacy against tuberculosis is driven by penetration into lesions and activity against resident bacterial populations. Antimicrob. Agents Chemother. *63*, e02516–e02518.

Sarathy, J.P., Via, L.E., Weiner, D., Blanc, L., Boshoff, H., Eugenin, E.A., Barry, C.E., and Dartois, V.A. (2018). Extreme drug tolerance of *Mycobacterium tuberculosis* in caseum. Antimicrob. Agents Chemother. *62*, e02266–17.

Szalai, B., Subramanian, V., Holland, C.H., Alföldi, R., Puskás, L.G., and Saez-Rodríguez, J. (2019). Signatures of cell death and proliferation in perturbation transcriptomics data—from confounding factor to effective prediction. Nucleic Acids Res. *47*, 10010–10026.

Tiwari, S., Tonder, A.J.van, Vilchèze, C., Mendes, V., Thomas, S.E., Malek, A., Chen, B., Chen, M., Kim, J., Blundell, T.L., et al. (2018). Arginine-deprivation–induced oxidative damage sterilizes *Mycobacterium tuberculosis*. PNAS *115*, 9779–9784.

Turkarslan, S., Peterson, E.J.R., Rustad, T.R., Minch, K.J., Reiss, D.J., Morrison, R., Ma, S., Price, N.D., Sherman, D.R., and Baliga, N.S. (2015). A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis*. Sci. Data *2*, 150010.

Vignali, M., Armour, C.D., Chen, J., Morrison, R., Castle, J.C., Biery, M.C., Bouzek, H., Moon, W., Babak, T., Fried, M., et al. (2011). NSR-seq transcriptional profiling enables identification of a gene signature of *Plasmodium falciparum* parasites infecting children. J. Clin. Invest. *121*, 1119–1129.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Zhang, Y., Permar, S., and Sun, Z. (2002). Conditions that may affect the results of susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. J. Med. Microbiol. *51*, 42–49.

![CellPress logo] **CellPress**
OPEN ACCESS

Cell Reports Methods
Article

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Middlebrook 7H9 Broth Base (7H9 broth)	Millipore-Sigma	M0178-500G
Middlebrook 7H10 Agar Base (7H10 agar)	Millipore-Sigma	M0303
Bedaquiline	Millipore-Sigma	843663-66-1
Clofazimine	Millipore-Sigma	2030-63-9
Isoniazid	Millipore-Sigma	54-85-3
Linezolid	Millipore-Sigma	165800-03-3
Moxifloxacin hydrocholoride	Millipore-Sigma	186826-86-8
Pretomanid	Millipore-Sigma	187235-37-6
Pyrazinecarboxylic acid (Pyrazinoic acid)	Millipore-Sigma	98-97-5
Rifampicin	Millipore-Sigma	13292-46-1
SuperScript II Reverse Transcriptase	ThermoFisher	18064014
Critical commercial assays		
Ribo-Zero Bacteria rRNA Removal Kit	Illumina	20040526
TruSeq Stranded mRNA HT library preparation kit	Illumina	20020595
Deposited data		
Transcriptomes from single drug treated Mtb	This paper	GEO: GSE165673
Mtb Transcriptome compendium for training of DRonA	This paper	GitHub: baliga-lab/DRonA_MLSynergy
Trained DRonA model (MTB_2020) used in this paper	This paper	GitHub: baliga-lab/DRonA_MLSynergy
Experimental models: Organisms/strains		
Mycobacterium tuberculosis: H37Rv	ATCC	27294
Software and algorithms		
Python	https://www.python.org/	N/A
SciPy	https://www.scipy.org/	N/A
GEOparser	This paper	Zenodo: https://doi.org/10.5281/zenodo.5598251, GitHub: baliga-lab/DRonA_MLSynergy
DRonA	This paper	Zenodo: https://doi.org/10.5281/zenodo.5598251, GitHub: baliga-lab/DRonA_MLSynergy
MLSynergy	This paper	Zenodo: https://doi.org/10.5281/zenodo.5598251, GitHub: baliga-lab/DRonA_MLSynergy
Google Colab notebook for DRonA and MLSynergy	This paper	Zenodo: https://doi.org/10.5281/zenodo.5598725, GitHub: baliga-lab/Google-colab-notebooks/blob/master/DRonA_MLSYnergy.ipynb

RESOURCE AVAILABILITY

Lead contact
Additional information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nitin S. Baliga (nitin.baliga@isbscience.org)

Materials availability
This study did not generate new unique materials nor reagents

Data and code availability

- The raw sequencing data have been deposited in GEO with accession number GSE165673. Information also listed in the key resources table.
- All original code for GEOparser, DRonA and MLSynergy has been deposited in Zenodo at https://doi.org/10.5281/zenodo.5598251, GitHub (baliga-lab/DRonA_MLSynergy), and MTB Network Portal (http://networks.systemsbiology.net/mtb/). Implementation of DRonA and MLSynergy on Google Colab server has been deposited in Zenodo at https://doi.org/10.5281/zenodo.5598725 and GitHub (baliga-lab/Google-colab-notebooks/blob/master/DRonA_MLSYnergy.ipynb). Information also listed in the key resources table.
- Any additional information required to analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains and growth conditions

Mycobacterium tuberculosis strains used in this study was H37Rv. Mtb cells were culture in standard 7H9-rich media consisting of 7H9 broth with 0.05% Tween-80, 0.2% glycerol, and 10% Middlebrook ADC. Frozen 1 mL stocks of Mtb cells were added to 7H9 medium and grown with mild agitation in a 37°C incubator until the culture reached an OD_{600} of ~0.4–0.8. The cells were then diluted to OD_{600} of 0.05 and added to 7H9-rich medium containing drugs at the predetermined amounts.

METHOD DETAILS

Minimum inhibitory concentration 50 (MIC_{50}) determination

10 mM working concentrations of drugs considered in the study were made with a suitable vehicle depending on drug solubility (*i.e.*, water, DMSO, or methanol). The 10 mM working concentrations of drugs were diluted in a twofold dilution series for 11 concentrations in 96-well plates. Each treatment series contained an untreated well as a control. Mtb H37Rv cultures were added to the wells and the plates were incubated at 37°C. Growth in cultures were measured as OD_{600} at 0 and 72 hours of incubation. All MIC_{50} determinations were performed in biological triplicate. Growth inhibition was determined by subtracting the initial reads from the final reads and then normalizing the data to no drug controls. Growth inhibition was fit to a sigmoidal curve and MIC_{50} was calculated for each drug (Table S3).

Time-kill assays

Using growth conditions described above, cells were diluted into 7H9-rich media containing drugs at predetermined amounts, along with vehicle controls (Table S3). Samples were taken after 0, 24 and 72 h, serially diluted and plated on 7H10 agar plates. All time-kill assays were performed in biological triplicate. Relative CFUs were calculated as log_{10} ratio of CFUs/ml of culture observed at start of treatment (T_0) and after drug treatment.

Collection, RNA extraction, and analysis of single-drug transcriptomes

Using growth conditions described above, cells were diluted into 7H9-rich media containing drugs at predetermined amounts, along with vehicle controls (Tables S2 and S3). Samples, in biological triplicates, were collected after 24 and 72 h. Samples were centrifuged at high speed for 5 min, supernatant was discarded and cell pellet was immediately flash frozen in liquid nitrogen. Cell pellets were stored at −80°C until bead beating in a FastPrep 120 homogenizer and RNA extraction was performed as previously described (Peterson et al., 2020). Total RNA was depleted of ribosomal RNA using the Ribo-Zero Bacteria rRNA Removal Kit (Illumina, San Diego, CA). Quality and purity of the mRNA was determined with 2100 Bioanalyzer (Agilent, Santa Clara, CA). Sequencing libraries were prepared with TrueSeq Stranded mRNA HT library preparation kit (Illumina, San Diego, CA). All samples were sequenced on the NextSeq sequencing instrument in a high output 150 v2 flow cell. Paired-end 75 bp reads were checked for technical artifacts using Illumina default quality filtering steps. Raw FASTQ read data were processed using the R package DuffyNGS as previously described (Vignali et al., 2011). Read counts were further analyzed with Kallisto (Bray et al., 2016) and RPKM values were calculated.

Collection and curation of Mtb-GEO dataset for training of DRonA

GEOParser was developed to download transcript profiles and metadata of drug-treated and untreated samples of Mtb-H37Rv from Gene Expression Omnibus (GEO). GEOparser collected median spot intensity from microarray samples and Reads Per Kilobase of transcript, per Million mapped reads (RPKM) from RNA-seq samples. The compendium dataset was curated by removing samples with low coverage (i.e., samples with <70% of annotated Mtb genes). The curated dataset was normalized by rank normalization with the "minimum" method with SciPy package (Virtanen et al., 2020).

Manual labeling of Mtb transcriptomes

Using the metadata collected by GEOParser (Table S1), transcriptomes were labelled as "viable" if the sample description stated that Mtb cultures were grown in optimal growth conditions (mid-log phase of growth in 7H9-rich media, incubated at 37°C with aeration)

and "non-viable" if the sample description stated that Mtb cultures were treated with >1x MIC_{50} drug for >12 hours. The remaining transcriptomes were labeled as "unclassified". Labels were saved as a comma separated value (.csv) file (Table S1).

Training and running DRonA

Rank, normalized transcriptomes along with the labels (Table S1) was provided to the SC-SVM classifier to start the iterative training of DRonA. Each iteration consisted of the following steps: (1) a single class support vector machine (SC-SVM) was trained on the training set i.e. transcriptomes labelled as "viable"; (2) the accuracy of the trained SC-SVM was calculated with Equation 1 using the test set i.e. transcriptomes labelled as "non-viable" initially and ones classified as "viable" through the iteration process;

$$Accuracy = \frac{True\ positive + True\ negatives}{True\ positive + True\ negatives + False\ positive + False\ negative} \qquad \text{(Equation 1)}$$

(3) assessment of the accuracy, (4) using the trained SC-SVM from (1), viability was predicted in transcriptomes labelled as "unclassified", and (5) newly predicted viable transcriptomes from unclassified set were moved to the training set. The iterative process was stopped when the accuracy of the classifier dropped below an accuracy threshold (85%) or when no new transcriptomes from the unclassified set were found to be viable. The cell viability scores (CVS) were calculated for samples as the weighted sum of gene expression ranks using the trained SC-SVM. CVSs were normalized by subtracting the score of a sample with maximum score observed in that experiment.

Inference of multi-drug transcriptomes (triangulation)

Transcriptomes of the Mtb cultures treated with multi-drug combinations at effective doses were predicted by triangulation with the single-drug treated transcriptomes and untreated control. Triangulation was called through 'triangulate' function in the MLSynergy algorithm, which collects transcriptomes of the drugs in combination (each profiled as single-drug) and untreated control and averages them with geometric mean. The inferred multi-drug transcriptomes were then returned to DRonA for CVS determination.

Calculation of MLSynergy scores for drug combinations

Expected CVSs were obtained from DRonA with the transcriptomes of the single-drug treatments that make up the drug combination and "expected CVS" was calculated by averaging the CVSs of single-drug treatments. The "predicted CVS" was obtained from DRonA with the inferred transcriptome of the drug combination. MLSynergy score were calculated as the ratio of expected CVS and predicted CVS. Further, MLSynergy scores were log normalized (base 2) in reference to the average of MLSynergy scores of same drug combinations that are considered to be additive in nature.

Comparison of INDIGO-MTB and MLSynergy predictions

Two INDIGO (Ma et al., 2019) models were retrained with default parameters. Model-1 was trained with the complete dataset (202 combinations and 46 drugs) and Model-2 was trained with partial dataset (98 combinations and 40 drugs) which was obtained after excluding combinations with bedaquiline, clofazimine, linezolid, moxifloxacin, pretomanid and pyrazinamide. Both models were tested on the combinations given in Table S6. Transcriptomes provided in Ma et al. were used as input for the INDIGO models. Transcriptomes generated in this study (summarized in Table S2) were used as input for the MLSynergy.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis reported in this article were performed with SciPy package (Virtanen et al., 2020) in Python. The p value from the Student's t test, sample mean and SEM were used as indicated in Figures 2, 3, 4, 5D, S3, and S4. Statistically non-significant (NS) were considered with p value > 0.05 and other qualifying p values were indicated accordingly * < 0.05, ** < 0.01, and ***< 0.001. The correlations reported in Figures 3 and S3 were calculated as the average correlation coefficient, r, from 100 iterations performed with 70% randomly selected data, r and p values were reported in the figures.